

## 基于4种机器学习算法的妊娠期中药“禁忌慎”判别

姜皓<sup>1,2</sup>, 张冰<sup>1,2\*</sup>, 张晓朦<sup>1,2</sup>, 林志健<sup>1,2</sup>

1. 北京中医药大学中药学院, 北京 102488

2. 北京中医药大学 中药药物警戒与合理用药研究中心, 北京 102488

**摘要:** **目的** 识别禁忌表述尚不明确中药中的妊娠期禁忌药, 并将其划分为禁用药、忌用药和慎用药, 为妊娠期妇女安全用药提供依据。 **方法** 以《中国药典临床用药须知》2015年版收录的666味中药为研究对象, 选取其中“禁忌慎”分类明确的药物, 基于神经网络、支持向量机、朴素贝叶斯和随机森林4种机器学习算法建模。根据交叉验证的受试者工作特征(receiver operating characteristic, ROC)曲线下面积(area under ROC curve, AUC)与F1分数(F1 score, F1)评价模型优劣, 筛选相对最优算法建立妊娠期中药“禁忌慎”判别模型, 并应用该模型预测禁忌表述尚不明确的药物。 **结果** 纳入224种药物建模, 模型由基于神经网络的中药禁忌与否判别模型(AUC=0.952, F1=0.885)、基于支持向量机的禁忌药禁用与否判别模型(AUC=0.912, F1=0.779)、基于朴素贝叶斯的禁忌药忌用与否判别模型(AUC=0.843, F1=0.333)和基于神经网络的禁忌药慎用与否判别模型(AUC=0.932, F1=0.877)构成。应用模型成功划分442种禁忌表述尚不明确的中药, 预测提示妊娠期禁忌药163种, 禁用药8种, 忌用药1种, 慎用药134种。 **结论** 构建的判别模型具有良好的稳健性与预测能力, 适用于妊娠期中药“禁忌慎”判别评估, 指导临床用药决策与实践, 提高临床合理用药水平。

**关键词:** 中药; 妊娠期; 机器学习; “禁忌慎”; 神经网络; 支持向量机; 朴素贝叶斯; 随机森林

**中图分类号:** R285.64 **文献标志码:** A **文章编号:** 0253-2670(2021)24-7596-10

**DOI:** 10.7501/j.issn.0253-2670.2021.24.022

## Discriminant of “prohibiting, avoiding or using caution” in traditional Chinese medicine during pregnancy based on four machine learning algorithms

JIANG Hao<sup>1,2</sup>, ZHANG Bing<sup>1,2</sup>, ZHANG Xiao-meng<sup>1,2</sup>, LIN Zhi-jian<sup>1,2</sup>

1. School of Chinese Materia Medica, Beijing University of Chinese Medicine, Beijing 102488, China

2. Center for Pharmacovigilance and Rational Use of Chinese Medicine, Beijing University of Chinese Medicine, Beijing 102488, China

**Abstract: Objective** To provide evidence for the safe medication of pregnant women, the identification of contraindications in traditional Chinese medicine (TCM) during pregnancy was not clear, and divided them into prohibited drugs, avoided drugs and cautious drugs. **Methods** A total of 666 TCMS collected in the 2015 edition of *Instructions for Clinical Medications* were selected as the research objects. The drugs with clear classification of “prohibiting, avoiding or using caution” were selected from the research objects, and four machine learning algorithms, neural network, support vector machine, naive Bayes and random forest, were used for modeling. According to the cross-validated area under receiver operating characteristic (ROC) curve (AUC) and F1 score (F1) to evaluate the pros and cons of the model, and to screen the relatively optimal machine learning algorithm to establish the “prohibiting, avoiding or using caution” discriminant mode of TCM during pregnancy, and the model was used to predict the drugs whose contraindication was not clear. **Results** A total of 224 kinds of drugs were included to establish the model. The models consisted of a neural network-based discrimination model for the contraindications of TCM (AUC=0.952, F1=0.885), a support vector machine-based discrimination model for the prohibited drugs of contraindications (AUC=0.912, F1=0.779), a naive Bayes-based discrimination model for the avoided drugs of contraindications (AUC=0.843, F1=0.333), and a neural network-based discrimination

收稿日期: 2021-07-21

基金项目: 国家中医药管理局-国家中医药领军人才支持计划“岐黄学者”项目(10400633210004); 第三批国家高层次人才特殊支持计划(万人计划)教学名师项目(2020063320001)

作者简介: 姜皓, 硕士研究生, 研究方向为中药药物警戒与合理用药。E-mail: jianghtcm@163.com

\*通信作者: 张冰, 教授, 主任医师, 博士生导师, 研究方向为中药药物警戒与合理用药、中药防治代谢性疾病等。E-mail: zhangbing6@263.net

model for the cautious drugs of contraindications (AUC=0.932, F1=0.877). The application model successfully divided 442 kinds of TCM whose contraindications were not clear. The prediction suggested that there were 163 kinds of contraindicated drugs during pregnancy, eight kinds of prohibited drugs, one kind of avoided drugs and 134 kinds of cautious drugs. **Conclusion** The discriminant models have good robustness and predictive ability. They are suitable for the discriminant evaluation of the “prohibiting, avoiding or using caution” in TCMs during pregnancy. They could guide the decision-making and practice of clinical medication, and improve the level of clinical rational medication.

**Key words:** traditional Chinese medicine; pregnancy; machine learning; “prohibiting, avoiding or using caution”; neural network; support vector machine; naive Bayes; random forest

由于大多数药物能够通过胎盘屏障直接或间接影响胎儿的生长发育, 妊娠期用药不仅需要考量母体的获益, 更须关注药物对胎儿可能造成的风险<sup>[1]</sup>。有调研显示, 妊娠期妇女用药具有普遍性<sup>[2]</sup>。中药“禁忌慎”是围绕风险与获益对禁忌药使用限制程度的划分(“禁” > “忌” > “慎”), 为临床常见用药警示语。因此, 在我国进入“三孩”政策时期, 特别是高龄产妇逐渐增加的时代背景下<sup>[3]</sup>, 研究妊娠期妇女用药安全, 识别妊娠期禁忌药, 并划分其“禁忌慎”, 具有重要的现实意义与临床价值。

本研究基于神经网络、支持向量机、朴素贝叶斯和随机森林 4 种机器学习算法, 建立妊娠期中药“禁忌慎”判别模型, 预测妊娠期禁忌表述尚不明确的药物, 以期为临床用药决策提供更为直观的用药依据。

## 1 资料与方法

### 1.1 研究对象

鉴于《中国药典临床用药须知》2015 年版(简称《临床用药须知》)面向临床遣药组方、记载中药品种数较多、药物信息较为全面, 将其收录的 666 种中药作为研究对象。

### 1.2 数据来源

围绕 666 种中药, 整合权威禁忌信息, 提取药性功效、药理毒理等药物基本信息。信息来源于①权威专著:《中国药典》2020 年版、《临床用药须知》2015 年版。②国家“十三五”规划中药学教材: 2016 年王建等主编的《临床中药学》; 2016 年周祯祥等主编的《中药学》; 2017 年钟赣生主编的《中药学》。

### 1.3 数据规约

功效依据《中国药典》2020 年版规约, 如利尿、利小便统一为利尿, 利水消肿拆分为利水、消肿等, 形成 112 种单元功效术语集; 药理依据 2018 年彭成等主编的《中药药理学》规约, 如抗癌、抗肿瘤统一为抗肿瘤, 调节子宫拆分为兴奋子宫和抑制子宫等, 形成 95 种单元药理作用术语集; 毒理依据第

3 版《药理毒理学》, 规约为 15 种毒理作用类别, 包括肝毒性、胃肠毒性、肾毒性、神经毒性、心血管毒性、血液毒性、呼吸毒性、内分泌毒性、免疫毒性、生殖毒性、致突变毒性、致癌毒性、皮肤毒性、细胞毒性、其他等。

### 1.4 基于机器学习建立妊娠期中药“禁忌慎”判别模型

**1.4.1 划分药物组别** 根据禁忌与否, 将药物分为禁忌药、宜用药与禁忌表述尚不明确药。其中禁忌药根据禁忌用语的不同, 又分为禁用药、忌用药与慎用药。宜用药的纳入须满足以下 2 个条件: ①国家卫生健康委员会公布的 110 种药食两用中药、《中药药物警戒》<sup>[4]</sup>中划分为妊娠期 A 级的药物或妊娠疾病适用药; ②目前尚无明确证据指出存在妊娠期风险的品种。

**1.4.2 筛选判别指标** 将药性功效、药理毒理各项数据作为候选指标。采用 SPSS 25.0 软件, 基于秩和检验中指标的  $P$  值筛选判别指标, 将  $P < 0.05$ , 即具有统计学差异的候选指标作为判别指标; 基于秩和检验中判别指标的秩平均值差比较该判别指标对各组判别影响的趋向性<sup>[5]</sup>。同时, 结合  $\chi^2$  检验中的  $\chi^2$  值评估指标在判别过程中的重要性,  $\chi^2$  值越高越重要。

(1) 筛选中药“禁忌慎”基础判别指标: 基于秩和检验对比禁忌药与宜用药, 筛选判别禁忌药的“禁忌慎”基础(禁忌与否)判别指标。在禁忌与否判别指标中, 秩平均值差为禁忌药的秩平均值减去宜用药的秩平均值。差值为正, 代表该指标可能对禁忌药的影响更大, 被视为危险指标; 差值为负, 代表该指标可能对宜用药的影响更大, 被视为保护指标。

(2) 筛选中药“禁忌慎”特征判别指标: 包括①基于秩和检验对比禁用与慎忌药(慎用药和忌用药), 筛选判别禁用药的禁忌药禁用与否判别指标; ②基于秩和检验对比忌用药与慎禁药(慎用药

和禁用药), 筛选判别忌用药的禁忌药忌用与否指标; ③基于秩和检验对比慎用组与忌禁药(忌用药和禁用药), 筛选判别慎用组的禁忌药慎用与否指标。在特征判别指标中, 秩平均值差计算方式以禁忌药禁用与否指标为例, 禁用药的秩平均值减去禁忌药的秩平均值, 若差值为正, 说明该指标可能对禁用药判别的影响更大; 若差值为负, 则表示该指标可能对慎用药判别的影响更大。

**1.4.3 建立判别模型** 中药“禁忌慎”判别模型由基础判别模型与特征判别模型组成。基础判别模型, 即中药禁忌与否判别模型, 用于判别是否为禁忌药; 特征判别模型由禁忌药禁用与否判别模型、忌用与否判别模型和慎用与否判别模型组成, 用于判别禁忌药类别(禁用药、忌用药或是慎用药)。借助 Orange 3.27.3 软件, 基于筛选出的判别指标, 结合影响中药安全的因素多而复杂、中药数据多为非线性特点, 选取数据复杂关系分析能力强的神经网络、适合非线性数据处理的支持向量机、对数据缺失值不敏感的朴素贝叶斯和可解释性好的随机森林 4 种较为适合中医药数据挖掘的机器学习算法<sup>[6-7]</sup>, 建立相应的判别模型, 并通过超参数调整优化模型。

**1.4.4 优选判别模型** 采用交叉验证法评估模型性能, 优选预测能力好、稳健性强的判别模型。交叉验证时, 依据建模样本数量大小, 分层折叠次数为 10 或 5。将受试者工作特征(receiver operating characteristic, ROC) 曲线下面积(area under ROC curve, AUC) 和 F1 分数(F1 score, F1) 分别作为

评估模型预测能力与稳健性的指标, 并通过 ROC 曲线图观察模型敏感性与特异性之间的关系<sup>[8]</sup>。其中, 模型的 AUC 在 0.5 以上才具有预测价值。在 0.5~0.7 时代表模型具有较低的预测能力, 在 0.7~0.9 时代表模型具有良好的预测能力, 大于 0.9 时代表模型具有优秀的预测能力。最终筛选出相对最优机器学习算法建立 4 个子模型, 共同构成妊娠期中药“禁忌慎”判别示例模型。

**1.4.5 预测示例药物** 由于 50% 是模型常用的分类阈值, 且考虑将用药风险尽可能最小化, 所以本研究将 50% 作为预测的分类阈值。将禁忌表述尚不明确药物作为预测示例药物导入妊娠期中药“禁忌慎”判别模型开展预测, 主要分为以下 3 步: 第 1 步, 预测中药“禁忌慎”基础(禁忌) 概率。将禁忌表述尚不明确药物的指标数据导入“禁忌慎”基础判别模型中, 得出药物作为妊娠期“禁忌慎”中药的基础概率, 即成为妊娠期禁忌药的禁忌概率。若禁忌概率 > 50%, 则视为妊娠期禁忌药。第 2 步, 预测中药“禁忌慎”的特征概率。将预测得出的妊娠禁忌药指标数据导入“禁忌慎”特征判别模型中, 得到药物作为妊娠期“禁忌慎”中药的特征概率。第 3 步, 计算药物禁用概率、忌用概率和慎用概率。禁忌概率乘以“禁忌慎”特征概率, 得出妊娠期中药的禁用概率、忌用概率或慎用概率。若概率大于 50%, 本研究视其为禁用药、忌用药或慎用药。若 3 种概率中存在二者及以上皆大于 50%, 则以概率最高者作为最终预测结果。预测流程见图 1。

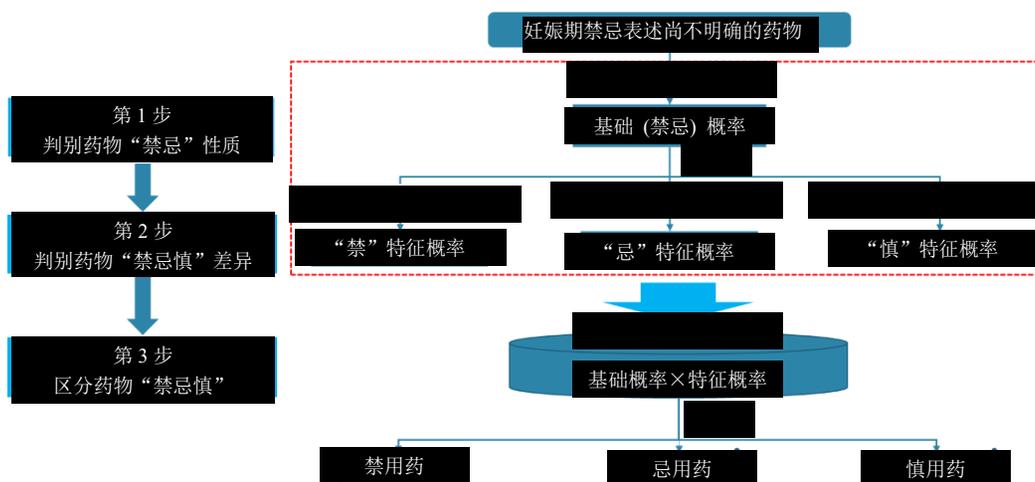


图 1 妊娠期中药“禁忌慎”预测流程图

Fig. 1 Flowchart for predicting “prohibiting, avoiding or using caution” of traditional Chinese medicine during pregnancy

2 结果

2.1 基本情况

纳入 224 种建模药物, 包括 120 种妊娠期禁忌药和 104 种宜用药。在妊娠期禁忌药中, 包括 48 种禁用、5 种忌用和 67 种慎用。纳入 442 种禁

忌表述尚不明确的药物, 即预测的示例药物。

2.2 妊娠期中药“禁忌慎”判别指标

在妊娠期中药“禁忌慎”基础(禁忌与否)判别指标中, 包括 36 项危险指标和 21 项保护指标, 分别见表 1、2。

表 1 妊娠期中药“禁忌慎”基础(禁忌与否)判别危险指标

Table 1 “Prohibiting, avoiding or using caution” basis (contraindication or not) discriminant risk index of TCM during pregnancy

类别	判别指标	秩平均值差	$\chi^2$	类别	判别指标	秩平均值差	$\chi^2$
五味	苦味**	39.18	30.02	功效	拔毒*	7.32	5.34
五味	辛味**	22.35	9.65	功效	温经*	5.60	5.20
四气	寒性*	14.54	9.28	功效	开窍*	6.39	4.51
毒性	毒性**	60.83	106.07	功效	散瘀*	8.76	4.24
归经	肝经**	24.12	5.17	功效	通乳*	4.67	4.33
归经	心包经*	4.67	4.33	药理	抗血栓*	9.41	3.83
功效	杀虫**	14.79	12.09	药理	扩张血管*	7.82	3.58
功效	止痛**	21.97	8.28	毒理	肝毒性**	25.78	14.06
功效	消痈肿**	20.03	9.28	毒理	肾毒性**	33.32	27.60
功效	活血**	15.44	10.31	毒理	神经毒性**	27.42	19.69
功效	逐水**	12.13	11.27	毒理	呼吸毒性**	20.31	10.53
功效	消疮**	9.33	8.67	毒理	心血管毒性**	35.97	28.60
功效	破血**	9.33	8.67	毒理	生殖毒性**	38.04	24.84
功效	散结**	13.14	6.36	毒理	皮肤毒性**	19.53	10.54
功效	逐瘀**	4.00	6.93	毒理	致癌毒性**	11.05	8.69
功效	通经络*	12.85	5.30	毒理	胃肠毒性**	12.49	7.06
功效	消癥*	6.53	6.07	毒理	血液毒性*	9.05	5.82
功效	泻下*	6.53	6.07	毒理	致突变毒性*	10.49	4.97

与妊娠期宜用药比较: \* $P < 0.05$  \*\* $P < 0.01$ , 表 1~5 同

\* $P < 0.05$  \*\* $P < 0.01$  vs appropriate medication during pregnancy, same as tables 1—5

表 2 妊娠期中药“禁忌慎”基础(禁忌与否)判别保护指标

Table 2 “Prohibiting, avoiding or using caution” basis (contraindication or not) discriminant protection index of TCM during pregnancy

类别	判别指标	秩平均值差	$\chi^2$	类别	判别指标	秩平均值差	$\chi^2$
五味	甘味**	-46.59	51.09	功效	健胃**	-7.54	8.08
四气	平性**	-16.44	5.72	功效	润肠**	-9.98	7.62
归经	肺经**	-22.62	6.61	功效	安胎**	-6.46	6.92
功效	补肾**	-13.07	12.13	功效	除烦**	-6.46	6.92
功效	健脾**	-18.45	17.83	功效	明目**	-7.69	6.52
功效	生津**	-13.07	12.13	功效	养肝**	-7.69	6.52
功效	补血**	-9.84	8.75	功效	温中**	-7.69	6.52
功效	滋阴**	-8.62	9.23	功效	固精*	-6.61	5.43
功效	补气**	-8.62	9.23	功效	固涩*	-5.53	4.34
功效	和胃**	-9.84	8.75	功效	止渴*	-5.53	4.34
功效	润肺**	-8.62	9.23				

在妊娠期中药“禁忌慎”特征判别指标中,包括33项禁忌药禁用与否判别指标、18项禁忌药忌用与否判别指标和31项禁忌药慎用与否判别指标,分别见表3~5。

### 2.3 妊娠期中药“禁忌慎”判别模型

完成模型的超参数调整后,根据模型性能表现,确立妊娠期中药“禁忌慎”判别模型的4个子模型,分别为基于神经网络的妊娠期中药“禁忌慎”基础(禁忌与否)判别模型、基于支持向量机的妊娠期禁忌药禁用与否判别模型、基于朴素贝叶斯的妊娠期

禁忌药忌用与否判别模型与基于神经网络的妊娠期禁忌药慎用与否判别模型。不同机器学习算法的最优超参数调整与性能指标情况分别见表6、7,ROC曲线图见图2。

### 2.4 妊娠期中药“禁忌慎”预测情况

**2.4.1 妊娠期禁忌药** 预测得出163种妊娠期禁忌药,见图3。

**2.4.2 妊娠期禁用药、忌用药和慎用药** 在163种妊娠期禁忌药,预测得出8种妊娠期禁用药,1种妊娠期忌用药,134种妊娠期慎用药,见图4。另

表3 妊娠期禁忌药禁用与否判别指标

Table 3 Discrimination indexes for whether prohibiting of taboo medicine during pregnancy

类别	判别指标	秩平均值差	$\chi^2$	类别	判别指标	秩平均值差	$\chi^2$
五味	甘味**	-16.23	22.39	功效	散寒*	-5.00	4.00
五味	辛味*	11.67	1.98	功效	通淋*	-5.00	4.00
四气	热性*	13.38	4.42	功效	解毒*	-8.33	3.33
毒性	毒性**	39.23	46.06	药理	抗肿瘤*	12.08	2.65
归经	胃经**	-14.58	5.83	药理	抑制胃液分泌*	3.75	4.50
归经	肾经*	10.83	4.17	药理	平喘*	-7.50	3.86
功效	清热**	-12.50	10.00	毒理	肾毒性**	20.42	10.53
功效	逐水**	12.08	10.78	毒理	生殖毒性**	20.83	8.33
功效	消疮**	8.33	6.57	毒理	心血管毒性**	23.33	12.44
功效	破血**	8.33	6.67	毒理	呼吸毒性**	15.83	7.76
功效	散瘀**	-6.58	6.30	毒理	肝毒性**	14.17	5.07
功效	消癥*	6.67	6.10	毒理	神经毒性**	13.33	5.02
功效	活血*	-10.42	5.21	毒理	皮肤毒性*	11.25	4.19
功效	拔毒*	7.08	5.35	毒理	内分泌毒性*	6.25	3.75
功效	补阳*	-5.83	4.67	毒理	胃肠毒性*	7.92	3.34
功效	解痉*	5.42	4.69	毒理	血液毒性*	6.67	3.56
功效	止血*	-5.00	4.00				

表4 妊娠期禁忌药忌用与否判别指标

Table 4 Discrimination indexes for whether avoiding of taboo medicine during pregnancy

类别	判别指标	秩平均值差	$\chi^2$	类别	判别指标	秩平均值差	$\chi^2$
五味	咸味*	18.78	4.70	药理	升血糖**	19.83	6.28
功效	温肾**	12.00	23.00	药理	扩张血管*	18.26	4.10
功效	消食**	14.48	10.52	药理	兴奋胃肠*	18.26	4.10
功效	活血**	27.13	5.88	药理	促进造血*	18.26	4.35
功效	退黄*	10.96	6.39	药理	抗应激*	18.26	4.35
功效	通经络*	25.57	4.54	药理	杀虫*	10.43	4.35
功效	排脓*	10.43	4.35	药理	收缩血管*	10.43	4.35
药理	抗肺损伤**	11.48	10.52	药理	保肝*	23.48	3.27
药理	保护肾脏**	20.35	7.35				

表5 妊娠期禁忌药慎用与否判别指标

Table 5 Discrimination indexes for whether using caution of taboo medicine during pregnancy

类别	判别指标	秩平均值差	$\chi^2$	类别	判别指标	秩平均值差	$\chi^2$
五味	甘味**	17.66	27.23	功效	拔毒*	-10.66	4.12
四气	热性*	-13.67	4.26	功效	凉血*	4.48	3.96
毒性	毒性**	-35.73	39.15	功效	通脉*	-7.27	3.96
归经	胃经**	13.09	4.83	功效	解痉*	-6.13	3.73
归经	膀胱经*	-7.92	3.96	药理	平喘*	8.49	5.07
归经	心包经*	4.48	3.96	药理	收缩血管*	-4.53	5.06
功效	清热**	13.43	11.87	药理	抗肿瘤*	-11.34	2.39
功效	逐水**	-7.27	8.63	药理	改善代谢*	4.48	3.96
功效	消癥**	4.48	8.85	毒理	心血管毒性**	-21.19	10.54
功效	散瘀*	8.49	5.07	毒理	肾毒性**	-16.66	7.21
功效	解毒*	9.80	4.74	毒理	呼吸毒性**	-14.81	6.99
功效	消疮*	5.37	5.21	毒理	生殖毒性**	-16.05	5.08
功效	破血*	-4.76	5.21	毒理	皮肤毒性*	-10.53	3.77
功效	化痰*	-9.58	4.31	毒理	肝毒性*	-10.58	2.90
功效	散寒*	0.43	4.75	毒理	神经毒性*	-10.10	2.96
功效	通淋*	5.37	4.75				

表6 妊娠期中药“禁忌慎”判别模型的最优超参数设置情况

Table 6 Optimal hyperparameter settings of “prohibiting, avoiding or using caution” discriminant model of TCM during pregnancy

机器算法	中药禁忌与否	禁忌药禁用与否	禁忌药忌用与否	禁忌药慎用与否
随机森林	树的数量设置为12;属性数小于4时不拆分	树的数量设置为12;属性数小于8时不拆分	树的数量设置为4;属性数小于9时不拆分	树的数量设置为9;属性数小于6时不拆分
支持向量机	常规 SVM;核为多项式, g=0.02, c=1.00, d=2.5;数值公差=0.001 0, 迭代极限为100	常规 SVM;核为多项式, g=0.01, c=1.00, d=2.5;数值公差=0.001 0, 迭代极限为150	常规 SVM;核为径向基函数, g=auto;数值公差=0.001 0, 迭代极限为100	常规 SVM;核为多项式, g=auto, c=1.00, d=2.5;数值公差=0.001 0, 迭代极限为100
神经网络	隐藏层中的神经元为100;激活函数为 ReLu;求解器为 L-BFGS-B;正则化, $\alpha=0.001$ ;最大迭代次数为150;可重复训练	隐藏层中的神经元为100;激活函数为 ReLu;求解器为 Adam;正则化, $\alpha=0.001$ ;最大迭代次数为100;可重复训练	隐藏层中的神经元为100;激活函数为 ReLu;求解器为 L-BFGS-B;正则化, $\alpha=0.001$ ;最大迭代次数为100;可重复训练	隐藏层中的神经元为100;激活函数为 tanh;求解器为 Adam;正则化, $\alpha=0.001$ ;最大迭代次数为100;可重复训练
朴素贝叶斯	无超参数调整			

表7 妊娠期中药“禁忌慎”判别模型评价指标比较表

Table 7 Comparison of evaluation indexes of “prohibiting, avoiding or using caution” discriminant model for TCM during pregnancy

机器算法	中药禁忌与否		禁忌药禁用与否		禁忌药忌用与否		禁忌药慎用与否	
	AUC	F1	AUC	F1	AUC	F1	AUC	F1
随机森林	0.939	0.888	0.851	0.779	0.600	0	0.884	0.806
支持向量机	0.947	0.886	0.912	0.779	0.896	0	0.912	0.851
神经网络	0.952	0.885	0.890	0.816	0.652	0	0.932	0.877
朴素贝叶斯	0.934	0.867	0.900	0.771	0.843	0.333	0.912	0.812

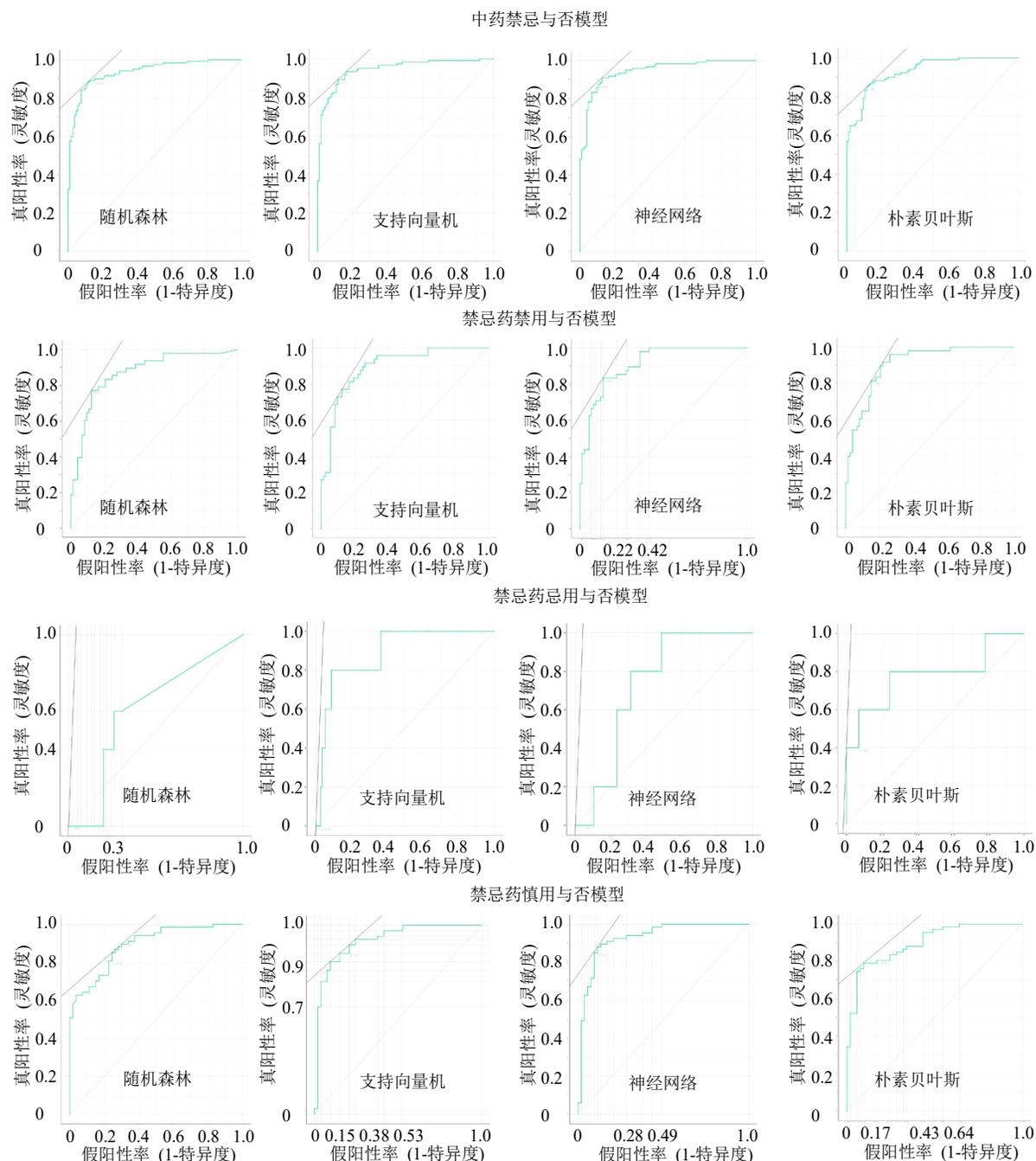


图2 妊娠期中药“禁忌慎”判别模型 ROC 曲线对比图

Fig. 2 Comparison of ROC curves of “prohibiting, avoiding or using caution” discriminant model for TCM during pregnancy

外, 尚有 20 种中药无法划分“禁忌慎”, 包括黄药子、凤仙透骨草、诃子、水红花子、石榴皮、儿茶、苍耳子、合欢皮、延胡索、前胡、暴马子皮、白屈菜、侧柏叶、苏合香、夏天无、补骨脂、老鹤草、青风藤、铁落、灯心草。

### 3 讨论

#### 3.1. 妊娠期中药“禁忌慎”判别指标的分析

禁忌药的判别基于基础判别指标中的危险指标

与保护指标。某味中药含有的危险性指标越多越重要, 则该药越有可能是禁忌药; 而含有的保护指标越多越重要, 则该药越有可能是宜用药, 即越不可能是禁忌药。

在已知某药为禁忌药的前提下, 筛选出“禁忌慎”各自的特征判别指标开展用药“禁忌慎”的判别。总的来说, 特征判别指标可分为 2 类: 第 1 类指标为“禁忌慎”中任意二者或 3 者所共有的判别指标;

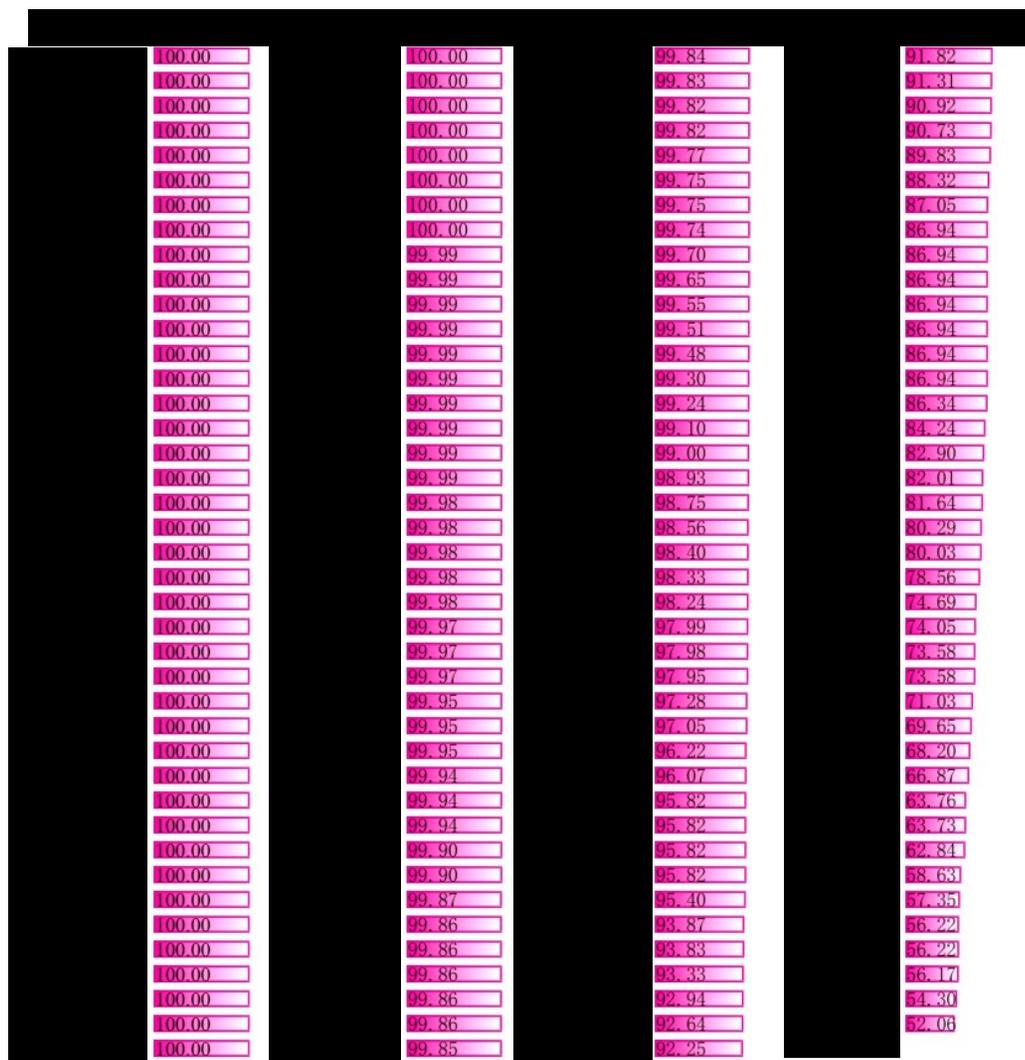


图3 妊娠禁忌药预测结果

Fig. 3 Prediction results of contraindications during pregnancy

第2类指标为“禁忌慎”各自独有的判别指标。

**3.1.1 基础判别指标对妊娠期中药“禁忌”性质的判别** 从危险指标来看,药性主要围绕毒性、苦味、辛味、寒性、肝经、心包经等;功效主要围绕杀虫、逐水、活血、消痈肿、止痛、消疮、破血、逐瘀、散结、消癥、泻下、通经络、拔毒、温经、开窍、通乳;药理主要围绕抗血栓与扩张血管;毒理主要围绕心血管毒性、肾毒性、生殖毒性、神经毒性、肝毒性、皮肤毒性、呼吸毒性、致癌毒性、胃肠毒性、血液毒性、致突变毒性等。毒性是药物偏性、峻烈程度及对人体毒副作用大小的高度概括。毒性越强,越有可能对母体或胎儿造成伤害。毒性与毒理密切相关,杀虫药多数具有一定毒性,散结、消癥及消痈肿往往药效较为峻烈,而辛香走窜、活血逐瘀或通利行经之品,有滑胎、堕胎之弊。因此,

在妊娠期服用与上述危险指标关联的中药会存在用药风险<sup>[9-10]</sup>。

从保护指标来看,药性主要围绕甘味、肺经、平性等;功效主要围绕健脾、补肾、生津、补血、滋阴、补气、和胃、润肺、健胃、安胎、除烦、明目、养肝、温中、固精、固涩、止渴等。甘味药具有补益、和中、调和药性和缓急止痛的功效。平性药往往药性平和、作用缓慢。甘平之药具有很好的安全性。因此,上述指标作为妊娠期禁忌药的保护指标,这是与临床实际妊娠期服药多为甘平的补益类中药、安胎类中药是相符的。

**3.1.2 特征判别指标对妊娠期禁忌药“禁忌慎”差异的确认** 从第1类指标来看,在妊娠期禁忌药中,具有辛味、热性、毒性,或存在逐水、拔毒、破血、解痉功效,或存在抗肿瘤药理作用,或存在肝毒性、

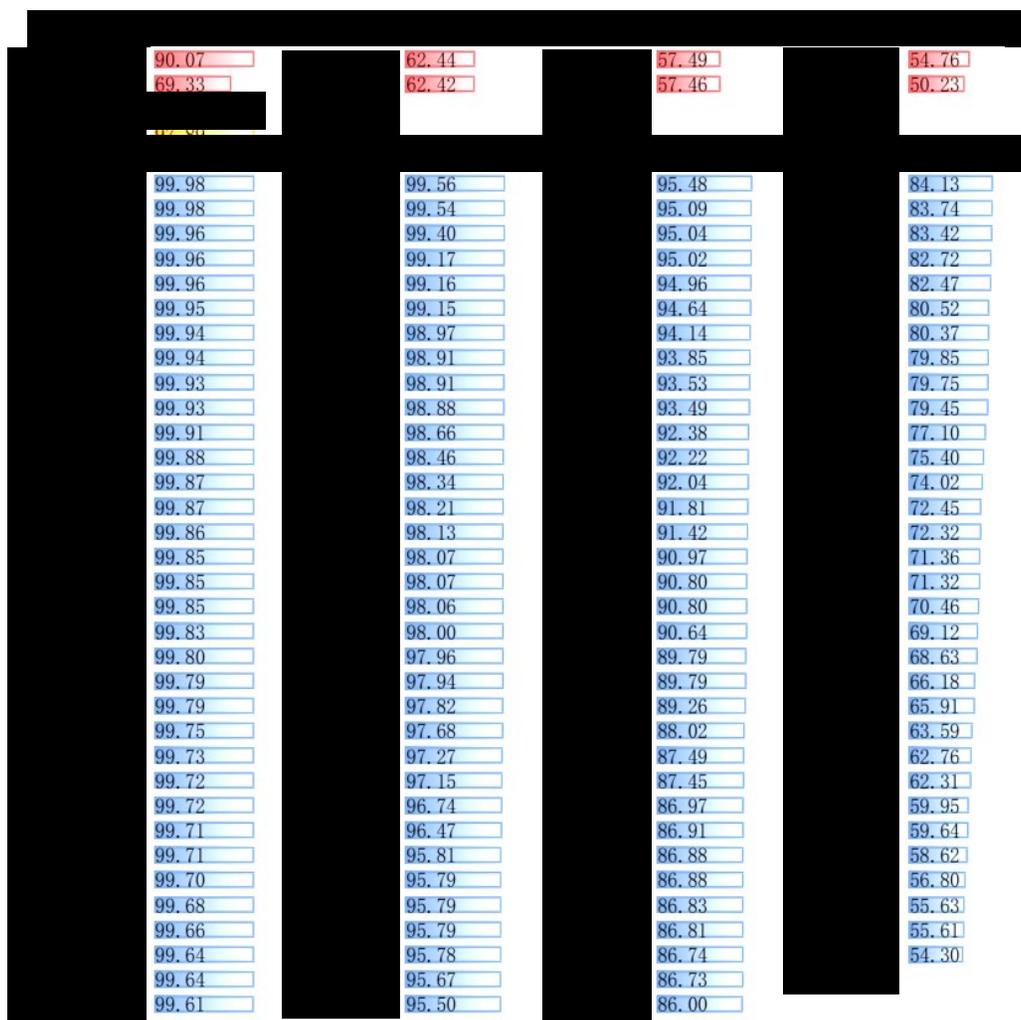


图 4 妊娠期禁用、忌用和慎用药物预测结果

Fig. 4 Prediction results about prohibited drugs, avoided drugs, and cautious drugs during pregnancy

肾毒性、神经毒性、呼吸毒性、心血管毒性、生殖毒性、皮肤毒性毒理作用的药物更可能为妊娠禁用、更不可能为妊娠慎用；具有甘味、归胃经，或存在散瘀、清热、解毒、散寒、通淋功效，或存在平喘药理作用的药物更可能为妊娠慎用，更不可能为妊娠禁用；存在消癥或消疮功效的药物更可能为妊娠禁用或妊娠慎用；存在活血功效的药物更可能为妊娠忌用，更不可能为妊娠禁用；存在收缩血管药理作用的药物更可能为妊娠忌用，更不可能为妊娠慎用。

从第 2 类指标来看，在妊娠禁忌药中，存在辛味、内分泌毒性、归肾经或抑制胃液分泌等判别指标，或不存在止血这一分类指标的药物更可能为妊娠禁用；存在扩张血管、消食、温肾、排脓、退黄、保肝、保护肾脏、促进造血、抗肺损伤、抗应激、杀虫、升血糖、调血脂、咸味、兴奋胃肠平滑

肌等判别指标的药物更可能为妊娠忌用；存在归心包经、改善代谢或凉血等判别指标，或不存在膀胱经、通脉或化痰等判别指标的药物更可能为妊娠期慎用。

### 3.2 妊娠期中药“禁忌慎”判别模型的预测能力比较

妊娠期中药“禁忌慎”判别模型识别妊娠期禁忌药及划分禁用与慎用的能力较好，但对妊娠期忌用预测的稳定性较差。针对妊娠期中药“禁忌慎”基础判别模型，其 AUC 和 F1 高于 0.850，说明该子模型具有良好的预测能力与稳健性，判别妊娠期禁忌药的结果可靠。在妊娠期“禁忌慎”特征判别模型中，妊娠期禁忌药禁用与否模型的 AUC 和 F1 高于 0.77，妊娠期禁忌药慎用与否判别模型的 AUC 与 F1 高于 0.87，说明此 2 个子模型亦具有较好的分类能力与稳健性能。妊娠期禁忌药忌用与

否判别模型的 AUC 虽然高于 0.84, 但 F1 仅为 0.33 说明该子模型的稳健性差, 无法较为精确识别妊娠期忌用药。

综合基础判别模型和特征判别模型分析模型对“禁忌慎”的判别能力: 从 AUC 来看, 划分“慎”优于“禁”优于“忌”, 表明模型对妊娠期慎用和非妊娠慎用药的区分效果最好; 从 F1 来看, 划分“慎”优于“禁”优于“忌”, 表明模型对妊娠期慎用药的预测过程中, 预测结果最稳定。可见, 妊娠期中药“禁忌慎”判别模型对慎用药的判别能力最强, 其次为禁用和忌用药。

### 3.3 妊娠期中药“禁忌慎”判别模型的局限性

本研究建立的模型主要存在两方面的局限性: 一方面从建模方法来看, 考虑到纳入数据量较少, 未设置测试集, 模型可能存在一定性能误估或过拟合风险; 另一方面从建模结果而言, 该模型对妊娠期忌用药的区分能力较差, 且部分可能的妊娠期禁忌药无法划分“禁忌慎”。造成以上局限性的最大原因在于基于有限数据难以实现模型理想化的“禁忌慎”判别或建模效果。在 3 种中药临床“禁忌慎”判别模型中, 部分子模型的样本量明显偏少, 如妊娠忌用药仅 5 种, 可能导致机器学习算法未能完全识别“禁忌慎”药物之间的差异性, 即无法获得建模所需的完整而准确的特征判别指标, 从而造成部分药物基于当前的模型时无法完成“禁忌慎”的判别评估。另外, 基础判别模型识别出的妊娠期禁忌药无法保证完全正确, 因此第 2 点局限性或许与部分无法区分“禁忌慎”的药物可能并非妊娠期禁忌药有关。

为改善以上局限性, 笔者提出 2 个方案: 一是扩大样本量, 增加作为研究对象的药味, 以提高“禁忌慎”分类明确药物的数量, 尤其是忌用药; 二是

判别指标的筛选结合专家咨询, 增强指标的可信度和合理性。

## 4 结语

中药妊娠期应用的安全性备受临床关注。本研究以 224 种“禁忌慎”分类明确药物为建模药物, 基于 4 种机器学习算法构建性能表现良好的妊娠期中药“禁忌慎”判别模型, 实现 442 种禁忌表述不明确药物的判别预测, 对临床具有较为直观的指导价值。今后有待以预测药物为研究对象, 设计毒理学实验、收集临床数据验证模型结果的可靠性, 进一步证明模型的可靠性与建模方法的可信度。

**利益冲突** 所有作者均声明不存在利益冲突

## 参考文献

- [1] 马晓玮, 顾红燕, 栗芳, 等. 我院妊娠期妇女门诊用药的合理性分析 [J]. 中国药房, 2021, 32(9): 1114-1117.
- [2] 张川, 张伶俐, 陈力, 等. 妊娠期用药调查研究的系统评价 [J]. 中国药学杂志, 2012, 47(11): 858-862.
- [3] 贾亮. 配套支持给力 三孩政策才能见成效 [J]. 北京观察, 2021(6): 30.
- [4] 张冰. 中药药物警戒 [M]. 北京: 人民卫生出版社, 2015: 112.
- [5] 兰欣, 卫荣, 蔡宏伟, 等. 机器学习算法在医疗领域中的应用 [J]. 医疗卫生装备, 2019, 40(3): 93-97.
- [6] 张恒, 张葆青. 基于数据挖掘的中医药临床研究述评 [J]. 中华中医药杂志, 2020, 35(11): 5378-5383.
- [7] 仇丽霞. 医学统计学 [M]. 第 3 版. 北京: 中国协和医科大学出版社, 2018: 426-448.
- [8] 周志华. 机器学习 [M]. 北京: 清华大学出版社, 2016: 33-35.
- [9] 徐建亚, 李思颖, 谢彤, 等. 基于羊水代谢组学策略评价妊娠禁忌中药的安全性 [J]. 南京中医药大学学报, 2019, 35(2): 223-228.
- [10] 郑依玲, 梅全喜, 戴卫波, 等. 妊娠禁忌中药研究概述 [J]. 中国药房, 2018, 29(3): 421-424.

[责任编辑 潘明佳]